ELSEVIER

Discrete Applied Mathematics 114 (2001) 155–169

DISCRETE APPLIED MATHEMATICS

On combinatorial properties of the Arshon sequence ☆

A.V. Klepinin*[,1], E.V. Sukhanov[1]

Department of Mechanics and Mathematics Ural State University, Lenin av., 51, 620083 Ekaterinburg, Russia

Abstract

We consider combinatorial and algebraic properties of the language of factors of the infinite sequence on the three-letter alphabet built by S.E. Arshon in 1930s. This sequence never contains two successive equal words, i. e., *avoids the number 2*. The notion of avoidability is extended from integers to rational numbers. It is shown that the avoidability bound for the considered language is $\frac{7}{4}$. This language is defined by two alternating morphisms; our method allows to study it like a formal language defined by one morphism. We also give a complete description of the syntactic congruence of the considered language.

Keywords: Arshon sequence; Pattern avoidance; Syntactic congruence

1. Introduction

Power-free words on finite alphabets play an essential part in the field of the formal language theory called combinatorics on words. In the beginning of XX century A. Thue built an infinite sequence on the two-letter alphabet which never contains three successive equal words. This sequence was rediscovered many times and was used in various combinatorial constructions (see, e.g., [11]).

Arshon [2] found a sequence on the three-letter alphabet which does not contain two successive equal words. This was precisely the sequence which was used as one of the basic elements for the solution of the bounded Burnside problem in group theory [1].

Other sequences with similar properties have been found later (see, e.g., [9,11,4]). Investigations on square-free sequences and their generalizations together with a series of related problems have led to exploration of languages defined by sets of forbidden words. These concepts were used in investigations on blocking (or complete) sets of words [7,8] and more algebraic blocking sets of terms and avoidable patterns [12,3].

☆ Translated from *Discrete* Anal. Oper. Res. (Novosibirsk) 6(2) (1999) 23–40.

* Corresponding author.

E-mail addresses: alklepin@mail.utnet.ru (A.V. Klepinin), evgeny.sukhanov@usu.ru (E.V. Sukhanov).

[1] Supported by the NSU grant center for mathematical research (Grant 46).

0166-218X/01/$ - see front matter
PII: S0166-218X(00)00368-1

Let v be a word on an alphabet Δ, and u be a word (a term, or a pattern) on the alphabet of variables Γ. The word v is said to contain the pattern u (or a value of the word u) if the word $h(u)$ occurs in v as a factor for some homeomorphism $h : \Gamma^+ \to \Delta^+$ with $h(x) \neq \emptyset$ for all $x \in \Gamma$. In this case, the word u is said to *block* the word v. Otherwise, if there are no values of u among factors of v, the word v is said to *avoid* u.

In these terms, any factor of the Thue sequence avoids the word $x^3 = xxx$, and factors of the Arshon sequence avoid the word x^2. This approach covers only "integer powers" of words. It is known [4] that the notion of avoidance can also be extended to "fractional powers" (for the definitions, see Section 3). Then the language of factors of the Thue sequence admits power 2 but avoids power $2+\varepsilon$ for all $\varepsilon > 0$. This means that Thue words are not only "cube-free", but satisfy a stronger property of being "strongly cube-free" [11].

Thus, the avoidability bound for the language of factors of the Thue word is equal to 2. The first goal of this paper is finding a similar bound for the language of factors of the Arshon sequence. This bound is equal to 7/4 and is given in Theorem 1. Another sequence on three symbols with the avoidability bound equal to 7/4 has been built in [6]. Our result demonstrates that the classical Arshon sequence can be used instead. The Arshon sequence is defined by iterating two alternating morphisms; at the same time, formal languages generated by iterating one morphism are more abundant. That is why the authors made an attempt to create a new method which would allow to deal with two morphisms in much the same way as with one morphism. The second main result of this paper is also obtained by this method.

It is well known that some interesting properties of a formal language can be derived from studying its syntactic semigroup, if non-empty words are considered. For example, some easy properties of the syntactic semigroup of the language of factors of the Arshon sequence show that this language is aperiodic and is not rational (the corresponding definitions can be found, e.g., in [10]). From the combinatorially algebraic point of view, a syntactic semigroup is the factor semigroup on the syntactic congruence of the language. The second main result of this paper is a description of the syntactic congruence of the language of factors of the Arshon sequence. This description is given by Theorem 2.

The structure of the paper is as follows. In Section 1 a summary of basic notions is given. Section 2 is devoted to the structural properties of the Arshon sequence following from its generating rule. Sections 3 and 4 are devoted, respectively, to finding the avoidability bound of the language of factors of the Arshon sequence and to studying its syntactic congruence.

2. Basic notions

Let us define the basic notions according to Salomaa [11]. An *alphabet* Σ is a finite non-empty set. The elements of Σ are called *letters* or *symbols*.

A *word* on the alphabet Σ is a finite string of letters of Σ. The string consisting of zero letters is called the *empty word* and is denoted by λ. The set of all words and of all non-empty words are denoted by Σ^* and Σ^+, respectively. In what follows, we shall mostly denote words by small Latin letters, although the Greek alphabet will be used for this purpose as well.

The *concatenation* of words x and y is the word xy. Concatenation is an associative operation with the empty word playing a part of the neutral element: $x\lambda = \lambda x = x$ for all x. Thus, Σ^+ is a semigroup and Σ^+ is a monoid. They are called *free* semigroup and monoid, respectively. For a word x and a positive integer i, the concatenation of i copies of x is denoted by x^i; by definition, we set $x^0 = \lambda$.

The *length* of a word x is denoted by $|x|$. By definition, we set $|\lambda| = 0$. It can be easily seen that

$$|xy| = |x| + |y|, \quad |x^i| = i|x|.$$

Let x and y be words, and let $y = x_1xx_2$ for some $x_1,x_2 \in \Sigma^*$. Then x is called a *factor* (or *subword*) of y. We shall denote this fact by $x \leqslant y$ (if $x \leqslant y$ and $x \neq y$, we shall write $x < y$) and say that x *occurs* in y. The *occurrence* of x is its concrete position in y. A *morphism* is a mapping h: $\Sigma^* \to \Delta^*$ satisfying $h(xy) = h(x)h(y)$; here Σ and Δ are alphabets. In other terms, h is a homomorphism of free monoids. An arbitrary subset of Σ^* is called a (*formal*) *language*.

Now, let us define the Arshon sequence according to Adian [1]. The main alphabet considered is $\Sigma = \{1,2,3\}$. The system of *Arshon words* $\{w_k\}_{k=1}^{\infty}$ will be defined by induction. By definition we set $w_1 = 1$. Let w_k be already built. Then w_{k+1} is obtained from w_k by substituting

$$\begin{array}{ll} \text{for symbols at odd positions :} & \text{for symbols at even positions} \\ 1 \to 123 & 1 \to 321 \\ 2 \to 231 & 2 \to 132 \\ 3 \to 312 & 3 \to 213. \end{array} \tag{1}$$

Thus, $w_2 = 123$, $w_3 = 123\,132\,312$, etc. We shall call the images of symbols *blocks*. The image of a letter from an odd (even) position will be called an odd (respectively, even) block. For example, the block 123 is odd, and the block 213 is even. Note that to know if the block is odd or even it is sufficient to know any two of its symbols; and if we know the parity of the block, then we can reconstruct it from any of its symbols.

Clearly, each of the words w_i is a concatenation of some blocks. Words having this property will be called *complete*, and the partition of a word to blocks will be called *natural*. To know the concrete position of a word v in w_i about the natural partition of w_i, we introduce two functions $\text{lsp}(v)$ and $\text{rsp}(v)$ as the number of letters not belonging to v in the left (respectively, right) block meeting v. Clearly, both functions can be equal only to 0, 1, or 2. The sense of there functions is illustrated by the following diagram: [2]

$$\cdots\star\star\star\ \ \star\star\star\ \ \underbrace{\underline{\star}}_{\mathrm{lsp}(v)}\overbrace{\underbrace{\star\star\ \ \star\cdots\star\ \ \star}_{|v|}}^{v}\underbrace{\underline{\star\star}}_{\mathrm{rsp}(v)}\ \ \star\star\star\cdots$$

Here $\mathrm{lsp}(v) = 1$ and $\mathrm{rsp}(v) = 2$.

It can be easily seen that w_k is a *prefix* of each of w_i for $i \geqslant k$. Thus, we can consider the limit of the word sequence $\{w_k\}$, that is the right infinite string such that $\{w_k\}_{k=1}^{\infty}$ is a set of its prefixes. Such infinite strings are called *superwords*; the superword generated by the system $\{w_k\}_{k=1}^{\infty}$ is the Arshon sequence; it is denoted by A_ω. Thus, A_ω looks as follows:

$$A_\omega = \underbrace{\overbrace{\underbrace{123}_{w_2}\ \ 132\ \ 312}^{w_3}\ 321\ \ 312\ \ 132\ \ 312\ \ 321\ \ 231}_{w_4}\ \ 213\ \ 231\ \ 321\ \ 312\ \ 321\ \ 231\ \ \cdots$$

Here are some properties of the superword A_ω.

Lemma 1. (1) *Each odd block in A_ω starts at an odd position, and each even block starts at an even position.*
(2) *The neighbor blocks of an odd block are even and vice versa.*
(3) *The sequence A_ω is square-free, i.e., does not contain factors of the form xx, where $x \in \Sigma^+$.*
(4) *Each subword of A_ω occurs in it infinite number of times.*
(5) *A word of length at most 3^k occurs in A_ω if and only if it occurs in w_{k+5}.*

Proof. Properties (1) and (2) are obvious. Property (3) was proved by Arshon [2].

Let us prove Property (4). Indeed, let $v < A_\omega$. Then $v \leqslant w_k$ for some [3] $k \in \mathbb{N}_0$. The word w_k is obtained from $w_1 = 1$ by applying the substitution (1) $k-1$ times. But it can be easily checked using Properties (1)–(3) that the word $w_1 = 1$ occurs in odd positions of A_ω infinite number of times. Consequently, all "images" of this word occur in A_ω infinite number of times.

Let us prove Property (5). Let $|v| \leqslant 3^k$. It is sufficient to prove that $v \leqslant A_\omega$ implies $v \leqslant w_{k+5}$. It is easy to understand that under these conditions $v \leqslant u$, where u is obtained from one letter by applying the substitution (1) $k+1$ times. It also can be directly checked that the word w_4 contains all the letters of Σ at both odd and even positions. Thus, $v \leqslant u \leqslant w_{k+1}$. Lemma 1 is proved. □

The set of all factors occurring in A_ω is denoted by LA, i.e., $LA = \{v \mid v \leqslant A_\omega\}$. In spite of the fact that the generating rule is easy, the sequence A_ω has a rather complicated

[2] From now on spaces indicate the natural partition to blocks.

[3] From now on $\mathbb{N}$ is the set of positive integers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and $\mathbb{Z}$ is the set of integers.

local structure. That is why before considering the properties of LA we are interesting in, in the next section we shall state and prove some auxiliary statements which help to clarify the local structure of A_ω.

3. The auxiliary construction

Let v and u be occurrences of arbitrary words in A_ω. If v and u start (finish) in blocks of the same parity, we say that v and u are *left* (*right*) *block synchronous*. If the occurrences are left and right block synchronous, we call them simply *block synchronous*. If the occurrences v and u start (finish) at the same position of (perhaps distinct) blocks of the same parity, we say that they are *left* (*right*) *synchronous*. Occurrences which are both left and right synchronous will be called simply *synchronous*.

For example, the following occurrences of words $u=123$ and $v=232$ are left block synchronous but are not right block synchronous:

$$\underline{123}\ 132\ 31\underline{2\ 32}1.$$

Note that the notions of being synchronous are defined not for words but for their occurrences. For example, these two occurrences of the word $v=231$ are neither left nor right block synchronous:

$$1\underline{23\ 1}3\underline{2\ 31}2.$$

However, sometimes the occurrences of the same word may be synchronous:

$$123\ 132\ \underline{312}\ 321\ \underline{312}.$$

The synchronization of distinct occurrences of the same word to A_ω turns out to be related to the structure of the sequence A_ω.

Lemma 2. *Let $v \in LA$ and $|v| \geqslant 8$. Then all the occurrences of v to A_ω are block synchronous.*

Proof. Let us consider all possible cases. Let v_1 and v_2 be distinct occurrences of the word $v=a_1a_2a_3a_4a_5\ldots$ to A_ω. Let us show that v_1 and v_2 are left block synchronous. To do this, let us list all the possible positions of v_1 and v_2 with respect to the natural partition.

(1) If $\mathrm{lsp}(v_1) \in \{0,1\}$ and $\mathrm{lsp}(v_2) \in \{0,1\}$, then the word a_1a_2 uniquely determines the parity of the blocks where v_1 and v_2 start.

(2) Let $\mathrm{lsp}(v_1)=1$ and $\mathrm{lsp}(v_2)=2$; i.e., let

$$v_1 = a_1a_2\underline{a_3a_4a_5}\ a_6\ldots,$$

$$v_2 = a_1\underline{a_2a_3a_4}a_5a_6\ldots.$$

The underlined blocks must have the same parity since they contain the common two-letter factor a_3a_4. But the parities of adjacent blocks in A_ω are always distinct, so the occurrences v_1 and v_2 are block synchronous.

(3) The same argument is valid for any $\mathrm{lsp}(v_1) \in \{1,2\}$ and $\mathrm{lsp}(v_2) \in \{1,2\}$.

(4) Let us prove that the situation with $\mathrm{lsp}(v_1)=0$ and $\mathrm{lsp}(v_2)=2$ (or, symmetrically, $\mathrm{lsp}(v_1)=2$ and $\mathrm{lsp}(v_2)=0$) contradicts the assertion of the lemma. Indeed, in this situation the positions of v_1 and v_2 with respect to the natural partition are

$$v_1 = a_1a_2a_3\ldots,$$

$$v_2 = a_1\underline{a_2a_3}\ldots$$

Note that since $a_1a_2a_3$ is a block, we have $a_1 \neq a_2 \neq a_3 \neq a_1$. Thus, the underlined prefix a_2a_3 of a block determines the whole block $a_2a_3a_1$, and we have $v = a_1a_2a_3a_1\ldots$ (i.e., $a_4 = a_1$). Thus,

$$v_1 = a_1a_2a_3\underline{a_1}\ldots,$$

$$v_2 = a_1a_2a_3a_1\ldots$$

Since the parities of adjacent blocks are distinct, the underlined symbol a_1 uniquely determines the remainders of the block: $a_5a_6 = a_3a_2$. Hence $v = a_1a_2a_3a_1a_3a_2\ldots$. Since $|v| \geqslant 8$, by similar arguments we obtain that $v = a_1a_2a_3a_1a_3a_2a_1a_2\ldots$, and its occurrences v_1 and v_2 are

$$v_1 = a_1a_2a_3a_1a_3a_2a_1a_2\ldots,$$

$$v_2 = \underline{a_1}a_2a_3a_1a_3a_2a_1\underline{a_2}\ldots.$$

Once more, the remainders of blocks are uniquely determined by the underlined symbols. So, the occurrence v_2 looks as

$$\underbrace{\overbrace{a_3a_2a_1a_2a_3a_1}^{p}\overbrace{a_3a_2a_1a_2a_3a_1}^{p}}_{v_2} < A_\omega.$$

But the Arshon sequence is square-free, so it cannot contain the square pp. We have obtained a contradiction.

Similarly, it can be proved that v_1 and v_2 are right block synchronous. Lemma 2 is proved. □

The next lemma is the main result of this section. It will allow us to introduce the notions of image and inverse image of sufficiently long factors of A_ω.

Lemma 3. *If $v \in LA$ and $|v| \geqslant 9$, then all the occurrences of v in A_ω are synchronous.*

Proof. It follows from the previous lemma that all the occurrences of v in A_ω are block synchronous. So, to prove that any two occurrences v_1 and v_2 of v are synchronous, it is sufficient to show that $\mathrm{lsp}(v_1) = \mathrm{lsp}(v_2)$ and $\mathrm{rsp}(v_1) = \mathrm{rsp}(v_2)$. Let us prove the first equality by disproving the following cases:

(a) $\mathrm{lsp}(v_1) = 0$ and $\mathrm{lsp}(v_2) = 1$ (i.e., v_1 starts with the first position of a block, and v_2 starts with the second one).

(b) $\mathrm{lsp}(v_1) = 0$ and $\mathrm{lsp}(v_2) = 2$ (i.e., v_1 starts with the first position of a block, and v_2 starts with the third one).

(c) $\mathrm{lsp}(v_1) = 1$ and $\mathrm{lsp}(v_2) = 2$ (i.e., v_1 starts with the second position of a block, and v_2 starts with the third one).

We shall restrict ourselves to the case (a) since it seems to be the most interesting. For the remaining cases, similar arguments are valid.

So, let v_1 start with the first symbol of a block, and v_2 start from the second one; both these block are odd or even due to Lemma 2. Repeating the arguments from the proof of Lemma 2, we obtain that v has the form $v = a_1a_2a_3a_2a_1a_3a_1a_2a_3\ldots$, and

$$v_1 = a_1a_2a_3\ a_2a_1a_3\ a_1a_2a_3\ldots,$$

$$v_2 = \underline{a_1a_2}\ a_3a_2a_1\ a_3a_1a_2\ \underline{a_3}\ldots.$$

Using parity, we once again see that the underlined words can be uniquely reconstructed to complete blocks, and thus

$$\underbrace{\overbrace{a_3a_1a_2\ a_3a_2a_1}^{p}\overbrace{a_3a_1a_2\ a_3a_2a_1}^{p}}_{v_2} < A_\omega.$$

Since the Arshon sequence is square-free, we have obtained a contradiction.

The remaining cases and right synchronization are considered similarly. Lemma 3 is proved. □

Note that the bounds for lengths in Lemmas 2 and 3 are strict since there exist occurrences of $v = 1231321$ ($|v| = 7$) which are not block synchronous and occurrences of $v = 12321312$ ($|v| = 8$) which are not synchronous.

Lemma 3 leads to a construction which is important for examining the Arshon sequence. Consider a word $v \in LA$ such that $|v| \geqslant 9$. Due to Lemma 3, all its occurrences are synchronous. In particular, all of them start at positions of the same parity (see Lemma 1). But it means that all the occurrences of v in A_ω have the same image under the substitution (1). Let us denote it by $\varphi(v)$ and call it the *natural image* of v with respect to the substitution (1). Clearly, the notion of the natural image is well-defined only for words of A_ω of length at least 9 since we cannot be sure that the image of a shorter word is unique. We shall use the notation $\varphi(v)$ only in cases where it is correct.

Analogously, let $v \in LA$, $|v| \geqslant 9$, and let all the occurrences of v start with the first letter of a block and finish with the last letter of a block (i.e., $\mathrm{lsp}(v) = \mathrm{rsp}(v) = 0$, and v is complete). Then the *natural inverse image* of v is uniquely defined as the word

$u < A_\omega$ such that v can be obtained from u by applying substitution (1). We use the notation $u = \varphi^{-1}(v)$.

For example, if $v = 312321312$, then

$$\varphi(v) = 3123212312132313213123212321231$$

and $u = \varphi^{-1}(v) = 313$. Clearly, applying (1) to $u = 313$ we can obtain v or $v' = 213123213$, $v' \neq v$. Hence in general, the existence of the natural inverse image $\varphi^{-1}(u)$ does not guarantee the existence of the natural image $\varphi(\varphi^{-1}(u))$. However, if the natural image $\varphi(u)$ of a word $u \in LA$ exists, then $\varphi^{-1}\varphi(u)) = u$.

Thus, the following lemma holds:

Lemma 4. *Every word $v \in LA$ of length $|v| \geqslant 9$ has the following properties*:
(1) *the natural image $\varphi(v)$ exists*;
(2) *if* $\mathrm{lsp}(v) = \mathrm{rsp}(v) = 0$, *then there is a natural inverse image $\varphi^{-1}(v)$*;
(3) $\varphi^{-1}(\varphi(v)) = v$;
(4) *if the natural inverse image $\varphi^{-1}(v)$ exists and $|\varphi^{-1}(v)| \geqslant 9$, then $\varphi(\varphi^{-1}(v)) = v$.*

We have shown that the substitution (1) behaves on sufficiently long words like an "invertible" morphism. This fact will help us in studying the language LA.

4. The avoidability bound of LA

It was mentioned in previous sections that the sequence A_ω avoids the word x^2, i.e., does not contain factors of the form xx. The general notion of avoidability is defined as follows. Let Δ and Γ be alphabets. A word $v \in \Delta^*$ is said to *contain* a word $u \in \Gamma^*$ (which is called also a *pattern*) if $\vartheta(u) \leqslant v$ for an appropriate non-erasing morphism ϑ: $\Gamma^* \to \Delta^*$. Otherwise v is said to *avoid u*.

The notion of avoidability can be modified as follows. Let $v \in \Delta^*$. A number p is called a *period* of v if the symbols $v[k]$ and $v[k+p]$ coincide for all $k \in \{1, 2, \ldots, |v| - p\}$. The *least period* of v is denoted by $\mathrm{per}(v)$, and the value $\exp(v) = |v|/\mathrm{per}(v)$ is called the *exponent* of v. The *inherent exponent* of v is the value $\mathrm{hexp}(v) = \max\{\exp(u) \mid u \leqslant v\}$. The inherent exponent is defined also for a superword (v_ω): $\mathrm{hexp}(v_\omega) = \sup\{\exp(u) \mid u \leqslant v_\omega, |u| < \infty\}$; the same definition is valid for an arbitrary language L: $\mathrm{hexp}(L) = \sup\{\exp(u) \mid u < v \in L\}$.

If the inherent exponent of a word (or a superword, or a language) is less than k, we say that it *avoids the exponent k*; otherwise, we say that it *contains the exponent k*. Thus, the inherent exponent partition the set of exponents to the subsets of avoidable and unavoidable exponents. The bound between these subsets will be called the *avoidability bound.*

Clearly, if k is rational, we can give an equivalent definition: a word $v \in \Delta^*$ contains the exponent k if and only if there exist a number $n \in \mathbb{N}_0$ and words x and y such that $|x| = a$, $|y| = b$, the word $(xy)^n x$ is a factor of v, and $k \leqslant n + a/(a+b)$. In what follows, this equivalent definition will be more convenient for us. We shall work in terms of the Arshon sequence A_ω rather than of the language LA. Since A_ω avoids squares, we have $1 \leqslant \mathrm{hexp}(LA) \leqslant 2$. Moreover, the following theorem holds:

Theorem 1. *The avoidability bound of LA is equal to* $\frac{7}{4}$; *it is attained at the word* $3121312 < A_\omega$.

Proof. It follows from the bound $1 \leqslant \mathrm{hexp}(LA) \leqslant 2$ that to study the avoidability bound of LA it is sufficient to study the occurrences of words of the form vuv to A_ω. That is why from now on we deal with a word $vuv < A_\omega$; we denote $a = |v|$ and $b = |u|$.

Note that the language LA does not avoid the exponent $\frac{7}{4}$ since the word $vuv = 3121312$ (where $v = 312$ and $u = 1$) contains it and belongs to LA. Thus, by the definition of inherent exponent, it is sufficient to examine only words v and u such that $a/(a+b) > \frac{3}{4}$. So, we may assume that $a > 3b$.

Furthermore, if a factor of A_ω of the form xyx is sufficiently long, then it turns out that there exists a shorter factor of A_ω whose exponent is not less.

Indeed, let $2a + b \geqslant 19$. Then it follows from the inequality $a > 3b$ that $\frac{7a}{3} \geqslant 19$. This implies $a \geqslant \frac{57}{7} > 8$. But since $a \in \mathbb{N}$, we have $a \geqslant 9$. Hence all the occurrences of v are synchronous due to Lemma 3, and so are the occurrences of the word vuv since its length is also at least 9. That is why the forthcoming arguments will be valid for any occurrence of vuv.

If v does not start with the first letter of a block ($\mathrm{lsp}(v) > 0$), we can extend it to the left to the beginning of the block, i.e., we can find a word q of length $|q| = \mathrm{lsp}(v)$ such that $qvuv < A_\omega$ and $\mathrm{lsp}(qvuv) = 0$. Furthermore, if $|u| \leqslant |q|$, then due to synchronization of the occurrences of v, we have $q = q'u$ for an appropriate $q' \in \Sigma^*$. But then $qvuv = q'\underline{uvuv} < A_\omega$, which contradicts the fact that A_ω avoids squares.

Hence $|u| > |q|$. Due to synchronization of the occurrences of v, we obtain that $u = u'q$ for an appropriate $u' \in \Sigma^+$. Consequently, $qvu'qv < A_\omega$, and

$$\frac{|qv|}{|qvu'|} = \frac{|q| + |v|}{|q| + |v| + |u'|} = \frac{|q| + a}{|v| + |u'q|} = \frac{|q| + a}{a+b} > \frac{a}{a+b}.$$

Thus, the exponent of the obtained word $qvu'qv$ is greater than the exponent of vuv. Therefore, in what follows, we assume that $\mathrm{lsp}(v) = 0$; similarly, we may assume that $\mathrm{rsp}(v) = 0$.

So, it is sufficient to consider only complete words v. Hence, the word vuv is also complete, and its natural inverse image $\tilde{v}\tilde{u}\tilde{v} = \varphi^{-1}(vuv)$ exists; here $|\tilde{v}| = \frac{1}{3}|v|$ and $|\tilde{u}| = \frac{1}{3}|u|$. Thus,

$$\frac{|\tilde{v}|}{|\tilde{v}\tilde{u}|} = \frac{\frac{1}{3}a}{\frac{1}{3}(a+b)} = \frac{a}{a+b},$$

i.e., there exists a three times as short word having the same exponent as vuv. Thus, for every word $vuv < A_\omega$ satisfying

$$|vuv| \geqslant 19 \quad \frac{|v|}{|vu|} > \frac{3}{4},$$

we can construct a word of length less than 19 having the same exponent. We have proved the following:

Lemma 5. *To find the avoidability bound of the language LA, it is sufficient to consider words of length at most* 18 *having the form vuv.*

Note that in the proof of the lemma a necessary condition for the lengths of v and u has been obtained. Hence, to find the inherent exponent of the Arshon sequence, it is sufficient to consider words $vuv \in \Sigma^*$ such that $3 \leqslant 3b < a \leqslant 8$ (here $a = |v|,\ b = |u|$). All the possible pairs of values of a and b are listed in the following table:

$a = \|v\|$	$b = \|u\|$	$\exp(vuv) = (2a+b)/(a+b)$
4	1	$\frac{9}{5}$
5	1	$\frac{11}{6}$
6	1	$\frac{13}{7}$
7	1	$\frac{15}{8}$
7	2	$\frac{16}{9}$
8	1	$\frac{17}{9}$
8	2	$\frac{18}{10}$

Due to Lemma 1, we must only check if the words of the form vuv having lengths $a = |v|$ and $b = |u|$ occur in w_9, that is, in the prefix of length 3^8 of the Arshon sequence. Checking this with the use of computer, we obtain the following result:

Lemma 6. *The avoidability bound of the language $LA|_{18} = \{v \mid v \in LA, |v| \leqslant 18\}$ is equal to* 7/4.

Thus, it follows from Lemma 5 that $\mathrm{hexp}(LA) = \mathrm{hexp}(LA|_{18})$, whereas due to Lemma 6 we have $\mathrm{hexp}(LA|_{18}) = \frac{7}{4}$. Thus, $\mathrm{hexp}(LA) = \frac{7}{4}$, which completes the proof of Theorem 1.

We have found the avoidability bound of the language LA and proved that it is attained. In particular, we can claim that if a word vuv is a factor of A_ω, then $|u| \geqslant \frac{1}{3}|v|$. Thus, any two occurrences of the same word must lie at some distance from each other, and this distance grows with the length of the word.

This result is close to that of [6], where another sequence on the three-letter alphabet having avoidability bound $\frac{7}{4}$ has been constructed. That sequence is generated by a morphism with the length of each image of letter equal to 19, whereas the images of letters under substitution (1) have length 3.

On the other hand, it is known [4] that there are no words of length greater than 39 avoiding exponent $\frac{7}{4}$; so this exponent is the minimal possible inherent exponent of a sequence on the three-letter alphabet.

5. The syntactic congruence of *LA*

The *syntactic congruence* σ_L of a language L on an alphabet Σ is defined as follows: $(u,v) \in \sigma_L$ if and only if for arbitrary $p,q \in \Sigma^*$, the words puq and pvq belong or do not belong to L simultaneously. The factor semigroup Σ^+/σ_L is called the syntactic semigroup of L. We are interested in σ_{LA}.

The set $I = \Sigma^+ \setminus LA$ is an ideal in the free semigroup Σ^+. The congruence corresponding to this ideal is defined by ρ_I. One of its classes is the ideal I, whereas all the other classes contain a single element each. The congruence ρ_I is the *Rees congruence* of the semigroup Σ^+ on the ideal I.

Theorem 2. *The equality holds* $\sigma_{LA} = \rho_I$.

Before proving the theorem, we shall mention several auxiliary statements. First, by definition, the language LA is a union of some congruence classes of ρ_I. It is known that the syntactic congruence of a language is the largest congruence having this property. Thus, $\rho_I \subset \sigma_{LA}$. Furthermore, if $(u,v) \in \sigma_{LA}$, then the definition of the syntactic congruence implies the following:
(a) $u \notin LA \Rightarrow v \notin LA \Rightarrow (u,v) \in \rho_I$;
(b) $u \in LA \Rightarrow v \in LA$.
Thus, to prove the equality of the congruences, it is sufficient to verify that if $u,v \in LA$ and $(u,v) \in \sigma_{LA}$, then $u = v$.

First, let us prove the equality of the lengths of u and v. We begin with the following easy

Lemma 7. *Let* $pvq \in LA$ *and* $puq \in LA$ *where* $|p| \geqslant 9$ *and* $|q| \geqslant 9$. *Then* $|v| = |u| + 6k$ *for some* $k \in \mathbb{Z}$.

Proof. Indeed, due to Lemma 3 all the occurrences of p and q are synchronous. Hence, all the occurrences of p finish and all the occurrences of q begin with the same position of blocks of the same parity. But since adjacent blocks have different parity in A_ω, and since v and u are situated between occurrences of p and q, we obtain the equality $|v| - |u| = 6k$ for some $k \in \mathbb{Z}$. The lemma is proved. □

Let us strengthen this result. To this end, we show that the length of a word is uniquely determined by sufficiently long contexts. Note that the next lemma is technical in nature, and the bounds appear in it in the form they will be used in the proof.

Lemma 8. *Suppose* $pvq \in LA$, $puq \in LA$, *and the conditions*
(1) $\max\{|v|,|u|\} \leqslant 6 \cdot 2^k$;
(2) $\min\{|p|,|q|\} \geqslant 9 \cdot 3^k + \sum_{i=0}^{k} 4 \cdot 3^i$
are fulfilled for some $k \in \mathbb{N}_0$. *Then* $|v| = |u|$.

Proof. Let us carry out the induction on k.

Base of induction. Let $k = 0$. Then $1 \leqslant |v| \leqslant 6$, $1 \leqslant |u| \leqslant 6$, $|p| \geqslant 9$, and $|q| \geqslant 9$. But under these restrictions, $|v| = |u|$ due to Lemma 7.

Induction step. Let the assertion of the lemma holds for all $k < n$ $(n \geqslant 1)$. Let us show that it holds for $k = n$. So, let $|v|, |u| \leqslant 6 \cdot 2^n$; in what follows, v and u will mean the particular occurrences to A_ω in contexts pvq and puq. Let $|p|$ and $|q|$ be at least $9 \cdot 3^n + \sum_{i=0}^{n} 4 \cdot 3^i$. Then since $n \geqslant 1$, we have $|p|, |q| \geqslant 9$. Hence due to Lemma 3 all the occurrences of p and q to A_ω are synchronous. It means that there exist words α, β, γ, δ of length at most two such that $p = \delta p_1 \alpha$ and $q = \beta q_1 \gamma$, where p_1 and q_1 are complete.

$$\cdots\star\star\star\ \overbrace{\star\star\star\underbrace{\star\cdots\star}_{p_1}}^{p}\underbrace{\overbrace{\star\star\star\ \star\cdots\star\ \star\star\star\star}^{v}}_{v_1}\overbrace{\underbrace{\star\cdots\star}_{q_1}\ \star\star}^{q}\star\ \star\star\star\ \cdots$$

Consider the words $v_1 = \alpha v \beta$ and $u_1 = \alpha u \beta$. Clearly, due to the choice of α, β, γ, and δ, the occurrences of v_1 and v_2 are also complete. Thus, the natural inverse images $p_2 = \varphi^{-1}(p_1)$, $q_2 = \varphi^{-1}(q_1)$, $v_2 = \varphi^{-1}(v_1)$, and $u_2 = \varphi^{-1}(u_1)$ are defined. Here $p_2 v_2 q_2 < A_\omega$, $p_2 u_2 q_2 < A_\omega$, and the estimates hold

$$|p_2| = \tfrac{1}{3}|p_1| \geqslant \tfrac{1}{3}(|p| - 4) \geqslant \tfrac{1}{3}\left(9 \cdot 3^n + \sum_{i=1}^{n} 4 \cdot 3^i\right) = 9 \cdot 3^{n-1} + \sum_{i=0}^{n-1} 4 \cdot 3^i.$$

Similarly,

$$|q_2| \geqslant 9 \cdot 3^{n-1} + \sum_{i=0}^{n-1} 4 \cdot 3^i.$$

On the other hand, for $n \geqslant 1$ we have

$$|v_2| = \tfrac{1}{3}|v_1| \leqslant \tfrac{1}{3}(|v| + 4) = \tfrac{1}{3}(6 \cdot 2^n + 4) \leqslant 6 \cdot 2^{n-1},$$

and similarly,

$$|u_2| \leqslant 6 \cdot 2^{n-1}.$$

But then $|v_2| = |u_2|$ by the induction hypothesis. Thus,

$$\begin{aligned} |v| &= |v_1| - |\alpha| - |\beta| = 3|v_2| - |\alpha| - |\beta| \\ &= 3|u_2| - |\alpha| - |\beta| = |u_1| - |\alpha| - |\beta| = |u|. \end{aligned}$$

The induction step and the lemma are proved. □

Corollary 1. *If $v, u \in LA$ and $(v, u) \in \sigma_{LA}$, then $|v| = |u|$.*

Proof. Indeed, we may choose $k \in \mathbb{N}_0$ such that $\max\{|v|, |u|\} \leqslant 6 \cdot 2^k$. Consider an occurrence of v to A_ω such that at least $9 \cdot 3^k + \sum_{i=0}^{k} 4 \cdot 3^i$ symbols precede it in A_ω; such an occurrence exists due to Lemma 3. Let p and q be the left and the right contexts of this occurrence of v, i.e., $pvq < A_\omega$, and the lengths of p and q are at least $9 \cdot 3^k + \sum_{i=0}^{k} 4 \cdot 3^i$. Such p and q exist because of the choice of the occurrence of

v and of the fact that A_ω is right infinite. But then the word puq belongs to LA by the definition of the syntactic congruence. Hence it follows from Lemma 8 that $|v| = |u|$. The corollary is proved. □

The following statement can be proved by sorting a reasonable number of cases (e.g., with the use of a computer):

Lemma 9. *Let $u, v \in LA$, $(u, v) \in \sigma_{LA}$, and $|u| = |v| = 9$. Then $u = v$.*

Proof. It is sufficient to check all the pairs (u, v) satisfying the conditions $u < A_\omega$, $v < A_\omega$, $u \neq v$, and $|u| = |v| = 9$. It was checked by a computer that for each of these pairs there exist the contexts p and q of length at most 5 such that exactly one of the words pvq and puq lies in LA. Due to Lemma 1, since $|pvq|$ and $|puq|$ are at most 19, to check this fact it is sufficient to look over w_{10}.

So, the assertion of the lemma can be checked in a finite number of steps. The lemma is proved. □

6. The proof of Theorem 2

As shown above, it is sufficient to prove that $u, v \in LA$ and $(u, v) \in \sigma_{LA}$ imply $u = v$. Due to Lemma 8, we may assume that u and v have equal lengths.

So, let $u, v \in LA$, $(u, v) \in \sigma_{LA}$, and $|u| = |v| = k$. Let us prove that $u = v$. To do so, we disprove the inequality $u \neq v$ by induction on k.

Base of induction. Consider the following cases.

(1) $k \leqslant 25$. Then there exist such words α and β that $|a| = \mathrm{lsp}(u)$, $|b| = \mathrm{rsp}(u)$, $|\alpha u \beta| = 27$, and all the occurrences of $\alpha u \beta$ to A_ω are complete. From the definition of the syntactic congruence, we have $(\alpha u \beta, \alpha v \beta) \in \sigma_{LA}$;

(2) $k = 26$ and there exists such letter a that all the occurrences either of au or of ua are complete. Respectively, either $(au, av) \in \sigma_{LA}$ or $(ua, va) \in \sigma_{LA}$;

(3) $k = 27$, and all the occurrences of u are complete.

Any of these three cases imply the existence of a pair $(u, v) \in \sigma_{LA}$ such that $u \neq v$, $|u| = |v| = 27$, and all the occurrences of u to A_ω are complete.

It is easily seen that all the occurrences of v to A_ω must have this same property. Indeed, it is sufficient to choose contexts of lengths greater then 9 for an occurrence of u. Then v must occur in A_ω in the same contexts, which are synchronous. By the same arguments, the words u and v are synchronous.

But this means that the natural inverse images $u' = \varphi^{-1}(u)$ and $v' = \varphi^{-1}(u)$ are defined. Moreover, due to Lemma 4 we have $\varphi(u') = u$ and $\varphi(v') = v$. If we now assume that $u \neq v$, then $u' \neq v'$. Since the lengths of u' and v' are equal to 9, we have $(u', v') \notin \sigma_{LA}$ due to Lemma 9. Thus, by the definition of syntactic congruence there exist such words p' and q' that exactly one of the words $p'u'q'$ and $p'v'q'$ belongs to LA. Without loss of generality, we assume that $p'u'q' \in LA$ and $p'v'q' \notin LA$.

Due to Lemma 4, the natural image $\varphi(p'u'q') = puq$ is defined, and $puq \in LA$. Here p and q are appropriate images of p' and q' with respect to the substitution (1). Let us show that $pvq \notin LA$. For if $pvq \in LA$, then due to the synchronization of u and v and the fact that p and q are concatenations of blocks by the construction, the natural inverse image $\varphi^{-1}(pvq)$ is defined. Here for these particular occurrences of p and q the equalities hold $\varphi^{-1}(p) = p'$ and $\varphi^{-1}(q) = q'$. Thus, $p'v'q' \in LA$ contradicting the choice of p' and q'.

So, there exist such contexts p and q that $puq \in LA$ and $pvq \notin LA$. This implies $(u,v) \notin \sigma_{LA}$, a contradiction. The base of induction is proved. □

Induction step. Consider the remaining cases of $k = 26$, $k = 27$, and $k > 27$; the arguments for these cases will be similar.

Let $(u,v) \in \sigma_{\mathrm{LA}}$ and $|u| = |v| = k = n$. Suppose that the assertion is proved for $k \leqslant n$. Let us show that the only possibility is the equality $u = v$. Suppose that, on the contrary, $u \neq v$. Repeating the arguments from the proof of the base of induction, we can show that there exist such words s and t ($|s| = \mathrm{lsp}(u)$, $|t| = \mathrm{rsp}(u)$) that $(sut, svt) \in \sigma_{LA}$ and the words sut and svt are complete. Note that $|sut| = |svt| > 27$ due to the chosen base of induction. Since $|sut| = |svt| > 9$, the natural inverse images $u' = \varphi^{-1}(sut)$ and $v' = \varphi^{-1}(svt)$ are defined with $u' \neq v'$ (because $u \neq v$), and

$$|u'| = |v'| = \tfrac{1}{3}|svt| > 9,$$

$$|u'| = |v'| = \tfrac{1}{3}|svt| \leqslant (\tfrac{k}{3} + \tfrac{4}{3}) \leqslant k - 1 \quad k \geqslant 4.$$

Thus, by the induction hypothesis, $(u', v') \notin \sigma_{LA}$. Similarly to the proof of the base of induction, we may assume that there exist such contexts p' and q' that $p'u'q' \in LA$ and $p'v'q' \notin LA$. Exactly as above, the natural image $\varphi(p'u'q') = psutq$ is defined, and $psvtq \notin LA$. So, $(psutq, psvtq) \notin \sigma_{LA}$, which contradicts the choice of u and v.

So, $(u,v) \in \sigma_{LA}$ implies $u = v$. The induction step is justified, and the theorem is proved.

7. Unlinked bibliography

[5,12]

Acknowledgements

We express our gratitude to the members of the *Discrete Mathematics* seminar at the Ural University for helpful discussions. We are also grateful to the referee whose comments aided to improve the text of the paper.

References

[1] S.I. Adian, The Burnside problem and identities in groups, Ergetnisse der Mathematik und ihrer Grenzgetiete, Vol. 95, Springer, Berlin, 1979.

[2] S.E. Arshon, A proof of the existence of infinite asymmetric sequences on n symbols, Mat. Sb. 2 (44) (4) (1937) 769–779 (in Russian).

[3] D.R. Bean, A. Ehrenfeucht, G.F. McNulty, Avoidable patterns in strings of symbols, Pac. J. Math. 85 (2) (1979) 261–294.

[4] C. Choffrut, J. Karhumäki, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 1, Springer, Berlin, 1997, pp. 329–438.

[5] M. Crochemore, P. Goralcik, Mutually avoiding ternary words of small exponent, Int. J. Algebra Comput. 1 (4) (1991) 407–410 (for further reading).

[6] F. Dejean, Sur un théorème de Thue, J. Comb. Theory. Ser. A 13 (1) (1972) 90–99.

[7] A.A. Evdokimov, Complete sets of words and their numerical properties, Medody diskr. analiza v issled. ekstrem. struktur 39 (1983) 7–19 (in Russian).

[8] A.A. Evdokimov, V.A. Krajnev, Problems on complete sets of words, Proceedings of XXII regional Popov conference, Novosibirsk, 1979, pp. 105–107 (in Russian).

[9] G. Lallement, Semigroups and Combinatorial Applications, Wiley, New York, 1979.

[10] J.-E. Pin, Syntactic semigroups, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 1, Springer, Berlin, 1997, pp. 679–746.

[11] A. Salomaa, Jewels of Formal Language Theory, Computer Science Press, Rockville, 1981.

[12] A.I. Zimin, Blocking sets of terms, Mat. Sb. 119 (1982) 363–375 (in Russian, English translation: Math. USSR Sbornik, 47(2) (1984) 353–364) (for further reading).